

第 1 章

什么是内存

内存是一片存储区域,当计算机在工作时,计算机的大脑——微处理器在这片区域临时存储数据。内存可存储你的想法,表达你的创意,做你要做的工作。计算机的内存越大,它能做的工作也就越多。

本章描述了位于计算机中的内存。我们将在以下几方面展开讨论:首先我们将随着“导游”去“参观”一下计算机内部的工作过程,然后讨论位、字节、千字节的概念,最后将描述一下各类 Intel 及其兼容的微处理器是如何控制内存的。通过本章内容,我们将对 PC 机中内存是如何工作的问题建立起一个基本的认识。

1.1 揭开计算机面纱

首先请将计算机工作台上零乱的物品收拾起来,腾出地方,因为将要在计算机内部开始一次旅行,看一看,拨弄拨弄,找出内存、扩充槽、微处理器在哪里。不必担心,这一切都不会产生什么问题。

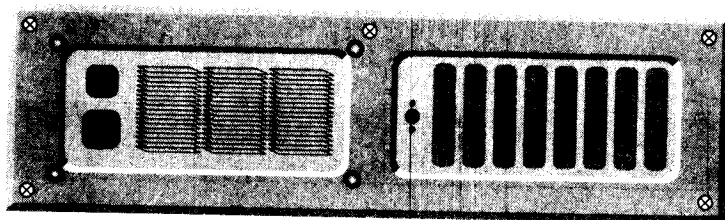
注意 不是每个人都是那么有能耐,操起一把螺丝刀,然后立即拔下他计算机的插头来观看一下他的计算机内部。但是如果你能够做的话,就应敢于上前把计算机盖打开。把办公室里的人都叫到身边来,勇敢一些!

1.1.1 打开机箱

在打开机箱前,应先将计算机电源关掉。接着还应立即把电源插头拔下来。(这一点是由教训而得出的经验。有一次我正在修理一台打印机,一位同事忽然发现打印机的电源插头拔下来了,他就尽职地将插头插上了,当时我的手还在打印机内!险不险?所以听我的话,关掉计算机,并且把电源插头拔下来。)把主机箱上的每一件东西(支架或系统的其他设备)都拿开。如果键盘插在支架的前面,就把它拔下来放在一边。

现在到主机的后面拧下用于将机壳固定在支架上的螺钉。也许需要把主机向外拉一点。将会发现有 2—6 颗螺钉,通常它们位于主机箱的角上、上部的中间,也许还有下部的中间。这些螺钉都是同一类型的,通常是十字头的螺钉。(不要拧下任何其他不同大小的螺钉。)参见图 1.1。

拧开这些螺钉后,拉着机箱盖向主机箱前滑动并且抬起一定的角度。(如果机箱盖不能移动,很可能是你忘记了拧开某个螺钉或箱盖卡在机箱前面的面板上了。)在拉开机箱盖时,要注意那些可能碰到的扁平电缆和电线。拉开时动作应慢一些,以便可以将挡路的



- ⊗ 拧开这些螺钉
- 不要拧开这些螺钉

图 1.1 拧开主机箱后面螺钉的示意图

电缆和电线挪开。打开机箱盖后,将它放在一边。

1.1.2 观察一下

每一种 PC 机都有不同的内部结构,但它们都有相似的部分。PC 机内部有通常由绿色玻璃纤维板制成的集成电路板,它上面爬满了电子“昆虫”,这就是主板。主板位于主机箱的底部。(参见图 1.2。)

在主板上找到内存芯片插排。它们可能位于组成几排或几列的一些小的“双列直插封装(dual in-line packages [DIPs])”的黑色 RAM 芯片(俗称“内存片”)中,或位于一些叫做“单片直插内存模块(single in-line memory modules [SIMMs])”的小插卡上(俗称“内存条”)。

有些主板,通常是基于 80386 或基于 i486 的计算机的主板上,有一些特殊的扩充插槽。它们不同于标准的扩充插槽(那些靠近计算机后部的插槽)。如果可能的话,就应在主板上找到它们。

另一些值得看一看的是电源(它包括风扇和为计算机其他部分供电的电源)、磁盘驱动器(包括硬盘驱动器和软盘驱动器)、扩充板和扩充卡、计算机的电池和扬声器(小喇叭)。

接下来找到微处理器。它是一个扁平的矩形或正方形的黑色芯片——像一块有小金属脚的 Keebler 巧克力。在芯片的上面可以看到一些数码如:8088、8086、80286、80386、i486、V20、V30,等等,如果有电线或电缆挡在那里,就轻轻地拨开它们。

注意 并不是每一个系统的微处理器都可以看到。例如在我的 Dell 386 中,微处理器位于硬盘的下面,它被挡住了,是看不见的。

最后,留心观看一下主板背面对应于扩充槽的那些位置,这些扩充槽是同计算机总线相连的。计算机总线是微处理器和其他设备的通信链路。专门设计的硬卡可以插入到这些扩充槽中。内存扩充卡(memory expansion card)就是这些专门设计的硬卡之一。当主板上已布满了内存芯片时,通过使用内存扩充卡可以使计算机增加更多的内存。

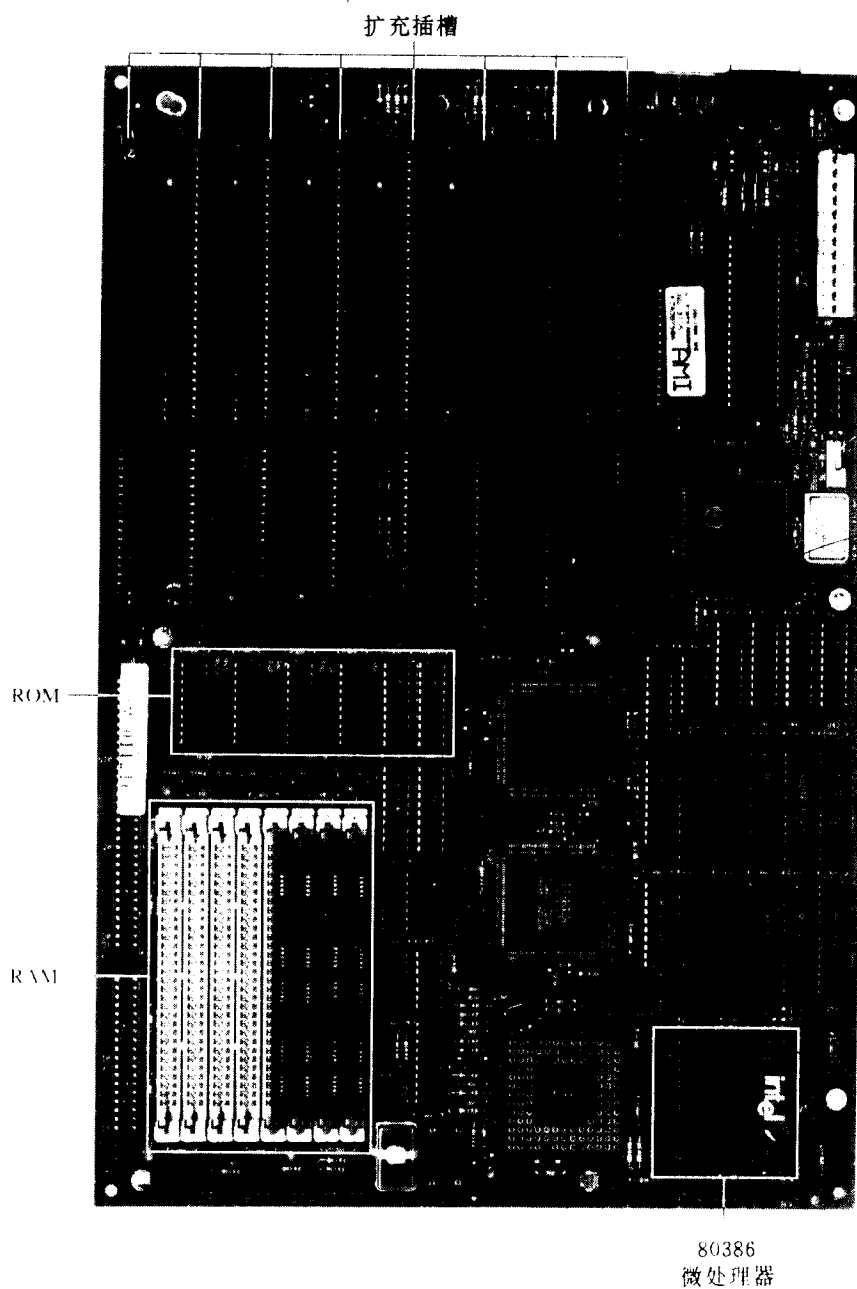


图 1.2 一个典型 PC 的主板

1.1.3 盖上机箱

现在结束在 PC 机中的旅行。把机箱盖放在主机箱的前上方,小心地把它向后滑动。注意不要缠住或碰到机箱内的电缆线。

注意 如果在此时插入了内存扩充卡,就不要马上将机箱盖盖上。最好的方法是开着机箱进行内存的测试。这样就可以排除错误,而不必反复打开和盖上机盖了。在插入扩充卡或进行排除错误的工作时,要注意将电源断开,千万不要在电源打开时移动扩充卡以及内存芯片。

把机箱盖尽量向后推。如果不能推到头,有可能是没有对齐;确认一下计算机的底面外设是否套在机盖里了。也可以检查一下计算机的锁是否被锁上了,并且检查一下盖子是否挂住了机器里的电缆。盖好后,将机盖上的螺钉拧上。现在将主机放回原位,连接上显示器(如果键盘已摘下,把它也连上),打开计算机电源。

1.2 表示内存的容量单位

内存是计算机存储信息的地方。但它又是如何将一首优美的诗或一幅有创意的 Escher 图之类的信息变成电信号而存储起来的呢?

计算机对词、声音、灰度或比例等一无所知。它能知道的仅仅是“开”和“关”——有电或没电。“开”对应于电压大约为 5V 的电流;“关”对应于没有电流。对我们来说,计算机在内存中用“开”和“关”来代表数值 1 和 0。

1 和 0 也许用起来不那么习惯,因为它们是二进制中的数字。或许因为我们人类有十个手指,所以我们使用十进制系统。十进制使用了数字 0—9。0 或 1 则可称为一个二进制数字(binary digit),可将此英文词缩写为“bit”(位)。

一位仅能存储 0 或 1 中的一个值。把多个位组合起来就可以增加可被存储的信息的数量。例如,两位可以存储 4 个二进制数:00、01、10 和 11。它们分别是十进制数的 0、1、2 和 3。如果你从前没有见过二进制数,这个二进制数序列看起来也许有些陌生。为什么 1 被存储为 01,但 2 却被存储为 10 呢?

现在来看一下十进制数 307。在这个数中的每一个数字都要乘上 10 的不同次幂:7 是个位数字(乘 10 的 0 次幂,任何数的零次幂是 1)(零的零次幂无意义——译者注);0 是十位数字(乘以 10 的 1 次幂);3 是百位数字(乘以 10 的 2 次幂)。这样十进制 307 就是 $3 \times 100 + 0 \times 10 + 7 \times 1$ 。

二进制也采用同样的原理来计数,只是每一个数字都要乘上 2 的不同次幂。让我们看一下二进制数字 101。最右边的数字乘以 2 的零次幂(1);它左边的数字乘以 2 的 1 次幂(2);再最左边的数字乘以 2 的 2 次幂(4);依此类推。因此,二进制数 101 就是 $1 \times 4 + 0 \times 2 + 1 \times 1$,它等于十进制数 5。

字节(byte)是一个由 8 位(bit)组成的位组。它是微处理器能够处理的位的最小集合。这一位组可以有 256 种不同的组合(通常,组合的数目可以通过求 2 的 n 次幂来得到,其中 n 是位组中的位数。2 的 8 次幂等于 256)。因此,一个字节可以存储的值的范围为 0—255。微处理器知道在内存中每一个字节的位置或地址,它可以存储或读出任意给定地址中的字节。

如果要想象一下内存中的字节,可以将它描绘成存储了一个信息的字符。例如单词

“door”需要内存中的 4 个字节；一段文字可能需要 400 个字节；一页文字，差不多需要 2000 个字节；依此类推下去，记录信息用的字节数就像掉在沙发下的甜饼上的灰尘一样在成倍地增长。

在计算机内添加了越来越多的内存时，就需要引进一个新的单位——千字节。千字节 (KB) 大约是 1000 个字节。(科学地讲，一个千字节是 1024 个字节，但是在二进制记数系统中，1024 是 2 的幂值中最接近于十进制数 1000 的数值。) 这样，一页文章需要大约 2 KB 的内存。

只需要很少的千字节即可存储文字内容。例如，在本书中，没有一个章节的文字内容需要多于 32 KB 的内存空间。(这本书总共也只占用了 316 KB。) 但是当你处理图像和声音的时候，就需要大量的内存了。

存储一幅图像所需要的字节数是一个不小的数字。观看一下你屏幕上显示的一幅玫瑰花图案，然后面对如下这个问题：这幅画使用了 500 KB 或更多的字节数！而同时一篇 60,000 个单词的文章只需要比这个数字还少的存储空间。所以我可以有把握地说，在计算机里，一幅画至少要顶上 1000 个单词。

在早期的个人计算机上，因为没有足够的内存，没有人梦想在 PC 机上存储彩色图像。但在今天，这已成为可能。最新型的个人计算机拥有上百万字节的内存。

兆字节 (MB) 大约是 1,000,000 字节。科学地讲，它是 1024 KB (1024 × 1024 字节或者说 1,048,576 字节，即 8,388,608 位——一个天文数字)。这并不是什么巨大的数目！一个 80386 微处理器可直接访问 4096 MB 的内存，或者说 4 千兆字节 (GB——gigabytes) 的内存。

内存的容量几乎关系到每一台计算机潜能的发挥。计算机拥有的内存越多，它能做的工作也就越多 (就像身处险境时你拥有的脑细胞越多，处理问题时也就越敏捷一样)。拥有更多的内存为存储不同类别的信息铺平了道路。我们可以存储展开的表格、音乐、图像等等。位 (bit)、字节 (byte)、千字节 (kilobytes)、兆字节 (megabytes)，这些术语一起描述了计算机所拥有的内存数量。

1.3 微处理器和内存

内存是必须要有的，因为计算机的大脑——微处理器 (CPU)，实际上并没有地方可用来存放信息。通常来说，拥有足够的内存是十分必要的。你也许想要内存容量越多越好。但实际上，微处理器的设计准确地限制了它可以使用的内存数量。

衡量微处理器功能好坏的其中一个指标，就是数据总线的宽度。数据总线像一条用于电信号行驶的高速公路，这条数据总线越宽 (就像高速公路上拥有更多的四车道)，在同一时刻，它能传输的数据也就越多。每个微处理器都有内部和外部两条数据总线。内部数据总线在微处理器内部各部件之间传输数据。外部数据总线负责微处理器和计算机其他部分 (如磁盘驱动器、扩充卡和内存) 之间的通信。

另一个衡量微处理器功能的指标是地址总线的宽度。地址总线宽度限定了微处理器可以直接访问的内存数量。例如，8086 拥有 20 条地址线，可直接寻址 1 MB 的内存；80286

拥有 24 条地址线,可直接寻址 16 MB 的内存。

微处理器的数据总线宽度通常用术语“位”来描述。例如,因为 8086 有一个 16 位的内部数据总线,所以称它为 16 位微处理器。在最早的 IBM PC 机中,所使用的微处理器就是 Intel 8088。8088 是一个“混血儿”,因为它有一条 8 位的外部数据总线和一条 16 位的内部数据总线,所以它被称为 8/16 位微处理器。

很让人奇怪的是,8088 的姐姐 8086,它的内部和外部数据总线都是 16 位的。但由于 8086 需要更多的外围芯片来支持,所以 IBM 公司最初便选择了 8088 作为 CPU。

忽略技术上的不同,8088 和 8086 都有一条 20 位的地址总线并且都可以直接寻址 1 MB 的内存。微处理器能够直接寻址的最大内存数量称为地址空间(address space)。

一个很重要的历史过程就是,所有的 PC 机都是“单性繁殖的”——包括现在放在你桌子上的计算机的系统,它们都是以原始的 PC 机为基础的。所有的计算机都遵循了相同的系统设计,这就使系统被 8088/8086 的 1 MB 地址空间所限制。其他一些设计因素也逐渐被考虑,这也是当今 PC 机内存结构繁杂的原因之一。(关于这一点,将在下一章“计算机使用内存的方式”中加以说明。)

PC/AT 采用了 80286 微处理器,准确地说,80286 是 8086 的孙子辈。(80186 是 Intel 公司继 8086 后的第二代微处理器,它也被用在一些计算机系统上,如著名的 Tandy 2000。但实际上这种微处理器没有提供比 8086 更突出的内存管理特征。)80286 的内部和外部数据总线都是 16 位的。它还拥有 24 位地址总线,使它可直接寻址 16 MB 的内存。

微处理器	位数	寻址范围(MB)
8088	8/16	1
8086	16	1
80286	16	16
80386SX	16/32	16
80386	32/32	4096
i486SX	32/32	4096
i486	32/32	4096
i486DX2	32/32	4096

80386 微处理器——PC 机梦寐以求的芯片,它拥有 32 位的内部和外部数据总线。它同 80286 以及 8086 是兼容的,并且拥有一条 32 位的地址线,所以它可以令人惊讶地寻址 4096 MB 的内存。

最精彩的是 80386 可以利用 640 KB 到 1 MB 之间的内存,使 MS-DOS 充分发挥其优势。因为这一功能,使得 80386 以及它的后代——i486,很快便成为 PC 机的标准微处理器。

80386 有一个年轻的姐姐——80386SX,它拥有一条 32 位的内部数据总线,同时又具有同 80386 完全相同的内存逻辑。不过 80386SX 仍被称为 16/32 位微处理器,因为它只有一条 16 位的外部数据总线。80386SX 的价格比 80386(也称为 80386DX)便宜,这使得于 80386SX 的 PC 机成为 80286 PC 机的理想替代品。

当前在 PC 机中使用的最流行的微处理器是 i486(包括 i486SX 和 i486DX2)。尽管它也拥有 32 位的数据总线和地址总线,并且提供了一些超过 80386 的功能和特征,但从内存的角度考虑,i486 和 80386 以及 80386SX 的工作方式完全相同。所以在本书中,术语 '386 代表任何基于 80386SX、80386、i486SX、i486、i486DX2 或 Pentium 微处理器的 PC 机。

注意 继 i486 之后的下一代微处理器是 Pentium。它可以被称为 586,尽管出于商标版权的考虑,Intel 公司倾向于将其命名为 Pentium。Pentium 同 80386 以及 i486 芯片兼容,因此本书中把 Pentium 和任何同它兼容的芯片并入到 '386 的家族中进行讨论。

通过回顾历史,可以使你了解到微处理器的功能是如何影响它所能使用的内存数量的。但是,问题在于 MS-DOS 是为 8088 以及它那有限的 1 MB 地址空间而编写的。MS-DOS 的早期版本把 '386 当做是一个快速的 8088。但这种方法并不十分有效,它没有充分发挥 '386 使用内存的潜能。这就是为什么 MS-DOS 6(以及版本 5)拥有不同于以前版本的优越性:它让你拥有大的内存空间,并且能够使用这一空间。

1.4 内存是如何工作的

如果你想刨根问底,一定要搞清一排扁平的芯片加上电流就能存储信息的道理,那么以下关于内存的说明将使你的好奇心得以满足。总的来说,计算机有两种基本的内存:RAM 和 ROM。

计算机的内存多数是随机访问存储器(RAM)。RAM 是一种计算机可读又可写的内存。RAM 是一种“不稳定”的存储器,当你关闭计算机时,存储在其中的任何信息都将丢失。有两种 RAM:动态 RAM 和静态 RAM(即 DRAM 和 SRAM)。

动态 RAM 具有大容量和低功耗的特点。它的存储单元是以一些存储电荷的电容为基础的,可以将它想象为一些极微小的电池。电池中有电荷或没有电荷分别代表逻辑上的 1 或 0。因为这些电池会渐渐失去它们的电荷,所以动态 RAM 需要定期地进行刷新——把一个个电子射进目标——以维持存储的数据。这种刷新所使用的周期降低了动态 RAM 的存取速度。

静态 RAM 使用与动态 RAM 不同的方法来存储 1 和 0,这种方法不需要刷新。静态 RAM 比动态 RAM 的速度要快,但它的容量小,并且制造成本高。出于这个原因,几乎所有的 RAM 芯片都是动态 RAM。

ROM 代表只读存储器(read-only memory),这种存储器不能写入或更新数据。ROM 是稳定的,即使在电源关上之后,ROM 芯片仍将保持它的内容。因此,ROM 被用来存储特殊的指令(如何在初始化的时候装入操作系统,如何控制硬件设备等)或其他计算机需要的基本程序代码。除非进入了一些非常重要的程序,否则这些 ROM 指令对用户来说是完全透明的,计算机完全不需要操作者介入,就可以找到和使用它们。

我们通过芯片能容纳的位数来评估某种 RAM 芯片。现在已有 16Kb(Kilobits)、

64Kb、128Kb、256Kb 和 1Mb 的 RAM 芯片。但是请注意,这些芯片存储的是单个的位,而不是将它们合并成字节。为什么呢?因为在 1 MB PC 及其兼容机中将 RAM 芯片排列成排(行)。因为一字节有 8 位,所以一排需要 8 个 RAM 芯片。然而在 PC 机上一排往往有 9 个芯片。每个芯片提供一个字节(8 位)中的一位,第 9 个芯片提供一个并行的位,这个并行的位提供了对其他 8 位的错误检查,以确认它们的可靠性,这一位便称为“校验位”。

注意 可以把一个 256Kb 的 RAM 芯片想象为一个 256,000 个单个位的栈。为了组成 256 KB,便需要 8 个(外加 1 个)这样的栈。这就是为什么一排 256 Kb(kilobits)的 RAM 芯片才等于 256 KB 的内存。对 1Mb 的芯片也是同样的道理——需要 9 片 1Mb 的芯片来组成 1 MB 的内存。

还有其他四个术语可帮助读者进一步理解内存。它们是访问时间、等待状态、交叉以及高速缓冲存储器。

1.4.1 访问时间

速度同容量一样也是衡量内存性能的一个重要指标。微处理器经常同内存打交道——读出或写入数据。例如,微处理器要将一个特殊的数存储在内存空间的 100,000 处。这需要经过内存存储这个数和完成刷新的时间。这一段延迟就称为内存的访问时间。访问时间是以纳秒(ns——百万分之一秒)来计算的。访问时间越短,内存的速度也就越快。

1.4.2 等待状态

理想的情况是,内存快到足以在微处理器准备存储其他数之前就完成数据的存入和刷新工作。如果不是这样,微处理器就必须等待 1 个或更多的时钟周期(clock cycles),等待在此期间内存完成刷新工作。(可以将“时间周期”想象成电子式的脉搏。)当内存在进行刷新时,微处理器必须等待的每一个时钟周期称为一个等待状态。零等待——一种最优化的安排,表明微处理器不再需要等待内存的刷新。

1.4.3 交叉(interleaving)

内存芯片传统上被排列在几行上。每一行都有一个内存地址范围,如图 1.3 所示。在

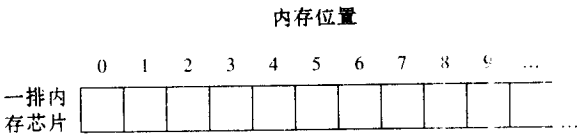


图 1.3 传统的内存芯片的安排

这种布局下,DRAM 芯片会跟不上今天 33MHz(或更快)的微处理器,因为它刷新时需要时间。静态 RAM 芯片的速度虽足够快,但它们的价格却非常昂贵。幸运的是,人们发明了

一种新的内存安排方案,它减小了慢速内存同快速微处理器之间的矛盾。这种管理方案被称为“交叉存储(interleaved memory)”。

典型情况下,程序是按顺序来访问内存的。如果某个程序要在内存位置 100,000 处存储一个数,该程序也许还想要将一个数存储在内存的 100,001 位置处,然后是 100,002,等等。因此交叉存储器将一个范围的内存位置分割在两排内存芯片中。第一排包括偶数地址的内存;第二排包括奇数地址的内存。如图 1.4 所示。各排芯片分别在奇数和偶数时钟周期时进行刷新。

现在,当微处理器在内存位置 100,000 处存储一个数的同时,内存位置 100,001 将被刷新。当微处理器在位置 100,001 处存储一个数时,内存位置 100,000 处将被刷新。这种简单的布置,帮助较慢的内存跟上了微处理器的速度。

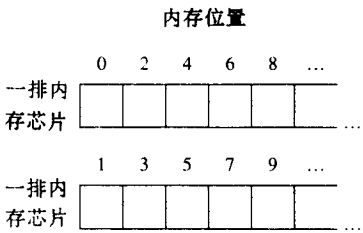


图 1.4 交叉存储内存芯片的安排

1.4.4 高速缓冲存储器

高速缓冲存储器(cache memory)是一个由极快的静态 RAM 构成的小容量的存储器(典型的大小是 8—256 KB),它位于微处理器和主内存之间。当微处理器向主内存中写入或从主内存中读出数据时,这个数据也被存储进高速缓冲存储器内。当微处理器再次需要这些数据时,微处理器将从高速缓冲存储器中读取这些数据,而不是访问较慢的主内存。Intel 公司直接将 8 KB 的高速缓冲存储器集成在 i486 微处理器中。

因为高速缓冲存储器使用静态 RAM,它所需要的成本,有时会使计算机的价格提高 300 美元左右。但是计算机运行速度的提高却是戏剧性的。如果你负担得起这笔增加的费用,那么最好是买一台带有高速缓冲存储器的计算机。高速缓冲存储器不同于在本书中讨论的其他类型的存储器,你不能自己为计算机增加高速缓冲存储器,因为高速缓冲存储器是主板上不可分离的一部分,所以必须买一台已经安装好高速缓冲存储器的系统。

注意 在购买扩充芯片前,要认真阅读一下计算机的用户手册。它会列出计算机所需要的内存芯片的规格和速度。不要为买快速但昂贵的芯片而感到心疼,如果使用较慢的芯片将引起不必要的等待状态,从而损害了系统的整体性能。

在本节中提到的术语或信息,对于理解 MS-DOS 是如何管理使用内存这个问题来说,并不是至关重要的。如果想知道更多的信息并且恰巧在美国的波士顿地区,那么波士顿的计算机博物馆将为你提供—个内存工作情况的优秀展览。不妨去参观—下这个博物馆。

1.5 小 结

内存是计算机用来临时存储数据的区域。计算机可以访问的内存数量和微处理器直

接相关。微处理器的功能越强——它的地址总线越宽,它能够访问的内存数量也就越多。早期版本的 MS-DOS 使设备驱动程序、内存驻留程序和应用程序不得不竞争 1 MB 中的低端 640 KB 内存的使用权,这 1 MB 的内存是原始的 PC 机微处理器可访问的最大空间。从 MS-DOS 5 以后,这种情况已经改变了。

- 位是一个二进制数字,它可以是 0 或 1 中的一个值。
- 八位组成一字节。
- 一个字节可以存储 256 个值中的任意一个数。可以把它看成是一个字符。
- 千字节大约是 1000 字节(科学地讲是 1024 字节)。
- 兆字节大约是 1,000,000 字节(科学地讲是 1,048,576 字节)。
- 8088/8086 微处理器分别有一个 8 位和 16 位的数据总线。它们可直接寻址 1 MB 的 RAM。
- 80286 微处理器有一个 16 位的数据总线,并且可直接寻址 16 MB 的 RAM。
- 80386 微处理器有一个 32 位的数据总线,并且可直接寻址 4096 MB 的 RAM。

第 2 章

计算机使用内存的方式

将计算机想象成全部都是 1 MB RAM,完全是理想主义的方式——在机器中有完全而连续的 1 MB 内存,整齐而又不出现断裂,就像一望无际、碧绿的高尔夫球场一样,没有一颗树,也没有任何沙坑。然而,实际上计算机中的内存并不是九片 RAM 芯片组成的无障碍且连续不中断的存储段。与之相反,内存之间是相互分开的,并根据机器的需要被分成几类。

本章介绍了典型 PC 机中内存的安排方式。此外,还描述了一些术语,到理解内存时,这些术语是至关重要的。

2.1 PC 机的内存

当 IBM 设计它的第一台微型计算机,即 PC 机时,它采用的是 8088 微处理器。8088 (以及 8086)能访问 1 MB RAM,也就是说,它的地址空间是 1 MB。

IBM 的工程师们必须将部分地址空间分配给必要的 ROM,而将其余的地址空间分配给 RAM,他们决定将低端的 640 KB 内存用作 RAM,上位(高端)的 384 KB 则保留给 ROM。PC 机首次亮相时,它并没有完整的一兆内存,然而界线却从此被确定下来了:低于 640 KB 为 RAM,供 MS-DOS 以及应用程序使用;高于 640 KB 则保留给 ROM、视频适配卡等使用。

- 常规内存即 PC 机的基本 RAM 区,从 0 KB 到 640 KB。常规内存也被称为低端 DOS 内存(low-DOS memory)。

- 上位内存指 640 KB 以上保留给 ROM 的内存区域。上位内存也可称作保留内存(reserved memory)或高端 DOS 内存(high-DOS memory)。

2.1.1 常规内存

常规内存是 MS-DOS 装入并运行用户程序的地方,其低地址部分被保留给计算机自身使用,从 2 KB 标记开始到 640 KB 的区域则用来运行应用程序。然而,这并不是说有整整 640 KB 的内存能全部提供给那些嗜好 RAM 的应用程序。毕竟,MS-DOS 要在常规内存中占一席之地,根据其版本号的不同,它要占用 18 KB 到 90 KB 的常规内存。

在 MS-DOS 之上的是数据存储区,MS-DOS 使用这些存储区来管理所有打开的文件,此外还有用 CONFIG.SYS 装入的所有设备驱动程序以及 AUTOEXEC.BAT 装入的内存驻留程序。最后在所有这些之上的才是用户的应用程序以及该应用程序用到的所有文件,例如某个文档,或者某个电子表格(spreadsheet)。

在最坏的情况下,如果已经装载了很多设备驱动程序和内存驻留程序,则可能只剩下

很少的几十兆字节的常规内存可用来运行应用程序。但是,这点内存够用吗?

2.1.2 上位内存区

IBM 将 PC 机内存的上端 384 KB 保留,用作未来的扩充,或用作 ROM。最初的 PC 机仅仅将该内存区的很小一部分分配给 BIOS,即基本输入/输出系统。BIOS 为 PC 机提供了控制磁盘驱动器、键盘等外部设备的较低级的指令。另一个小区域则被用作视频内存,用来存储单色显示器或彩色显示器显示的信息。此外,剩下的大部分空间仍等待着扩充。

塞满的 RAM

内存驻留程序的生产在八十年代中期达到顶峰。当时,几乎每一个程序都能将自己驻留,按下一个“热键”就能方便地对其访问。尽管很少有几个程序是确实需要被驻留的,不过将程序编制为内存驻留方式,确实是一种推销的策略。

内存驻留程序(也叫“终止并常驻”程序,即 TSR)带来了很多问题。首先,未确定一个编制 TSR 程序的标准,程序间常常发生冲突,每个程序都想抢夺对计算机的控制。此外,MS-DOS 对众多的内存驻留软件既不帮忙,也没有提供任何规则或约定。最后,内存驻留程序占用了大片常规内存区域,造成 RAM 拥挤(RAM cram)。

最终,TSR 的疯狂已渐渐平息。现在仍能找到一些内存驻留程序,不过绝大多数都是工具软件。人们编制“弹出式应用软件”的强烈欲望逐渐消逝,随着 Microsoft Windows 等多任务环境的出现,对内存驻留软件的需求较之以前已大为减少。

至今,上位内存区仍然没有被ROM填满。上位内存的前128KB被用作视频内存区(RAM),供大多数PC机中的单显、CGA、EGA或VGA图形系统使用;紧邻着的128KB则被保留给视频ROM、硬盘控制器ROM等可安装的ROM;最后的128KB被保留给ROMBIOS使用。图2.1采用传统的内存映象形式,给出了所有这些内存区的位置。

2.1.3 640 KB——MS-DOS 的樊篱

随着程序大小的不断增加,以及能访问更大内存的新型微处理器的相继出现,最初的PC机设计的局限性变得越来越明显。640 KB分界点不再仅仅被人们称之为程序RAM的终止点和上位内存开始点,有人将它称之为MS-DOS的樊篱——就像竖起了一堵墙,墙后面的内存使用被禁止起来。

程序只能在MS-DOS提供的640 KB或更少的RAM空间中工作。在1981年,这是当时任何其他微型计算机可用内存容量的十倍。但是,到八十年代末期时,这已经是被人们认为小得可怜的内存了。幸运的是,现在已经找到了解救这一困境的方法。

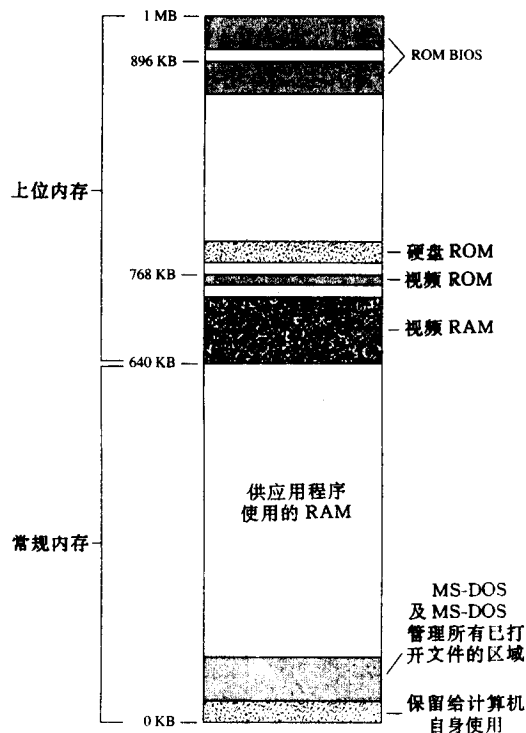


图 2.1 PC 机的内存图

2.2 扩充(expanded)内存

电子表格软件的用户称得上是使用大内存的第一大户。在用电子表格来工作时,表格存储在内存中;如果内存越大,则可生成的表格也就越大。对于家庭预算来说,256 KB 内存就足够了;小办公室中,640 KB 也是一个比较合适的内存;但对于大的表格用户来说,640 KB 内存就不够用了。

有人用硬件和软件结合起来设计了一个方案,用以解决内存不够的问题。Lotus、Intel 以及 Microsoft 公司专门联合设计了 LIM 扩充内存规范(EMS)标准,来管理扩充内存这一基本上可称之为 PC 机辅助内存的区域。扩充内存并不是 1 MB 标号以后的内存区域,应用程序也不能在扩充内存中运行。相反,它更像一个内存缓冲池,一个位于机器内部与 EMS 兼容的扩充卡中的内存存储区。与 EMS 兼容的软件能够访问到卡上的这块内存,显然,这意味着机器提供了更大的内存来存储数据。

为了访问和使用计算机上的扩充内存,用户必须有一个与 EMS 兼容的内存扩充卡和一个被称之为扩充内存管理器(EMM)的设备驱动程序,此外,当然也得有一个能够使用扩充内存的 EMS 兼容应用软件。

2.2.1 扩充内存的工作方式

扩充内存使用了上位内存的一个未用区域,它将上位内存中一个 64 KB 的内存块用作页面(page frame),参见图 2.2。PC 机中的扩充内存跟主存脱离开来,它独处一方,位于某个 PC 扩充槽中的扩充卡中。

EMM 设备驱动程序以四个 16 KB 大小的页映象到页面中的方式工作,从而让应用程序能够存取扩充内存。扩充内存由硬件和软件进行控制,需要时,16 KB 大小的页能被换进页面,或从页面中换出。存储页只要位于页面中,处理器就能访问该存储页,因为此时它已处于 1 MB 地址空间之内。

只要调用 EMM 设备驱动程序提供的各种功能,应用程序便可以将扩充内存的其他页映象到页面中,以上所述的过程就是扩充内存的使用方式。

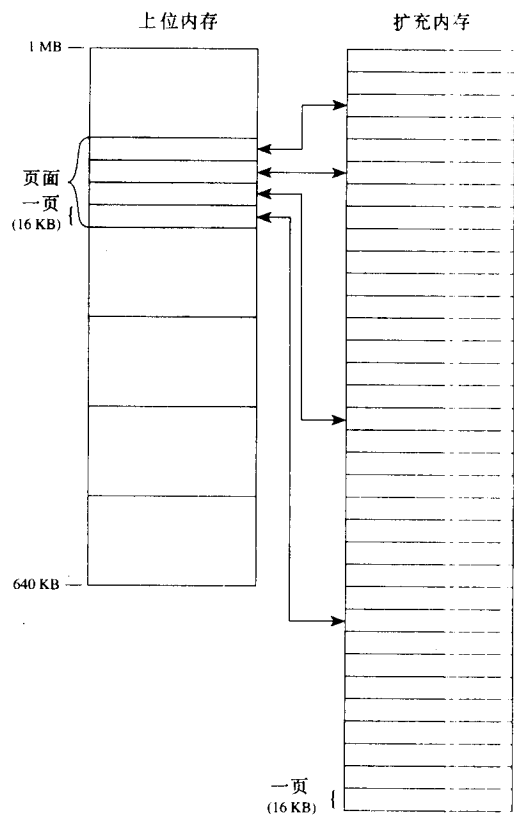


图 2.2 访问扩充内存的方式

注意 EMS 3.2 版实际上是专门为存储器设计的,而不是为运行程序设计的。将扩充内存中 16 KB 大小的页映象到页面中,对于保证快速访问电子表格的未用部分来说,确实不愧为一个好的方法。但是真正的程序都要求使用能被处理器直接访问到的连续内存。

EMS 4.0 版是 EMS 标准的最新版本,可支持对多达 32 MB 扩充内存的访问。EMS 4.0 标准允许新的 EMS 卡将程序和数据一样移进扩充内存。这样,扩充内存对于多任务处理变得更加有用。不过,它仍然要比直接寻址常规内存慢得多。

2.2.2 使用回填(backfill)

早期的 PC 机主板上仅装有 256 KB 甚至更少的内存,装一块内存扩充卡是其增加内存容量的唯一方式。随着内存价格的不断下降,目前大多数计算机都带有 1 MB 或更大容量的内存。

EMS 扩充内存卡早期的一个功能,除了提供扩充内存外,还能提供常规内存。EMS 卡可以为只配置有 256 KB 内存的 PC 机提供附加的 384 KB 常规内存,以组成一个完整的 640 KB 常规内存区。此外剩下的所有内存则被配置成扩充内存。这种用扩充内存填充常规内存的过程被称之为“回填”。

如果你的 8088、8086 或 80286 机器已经有 640 KB 常规内存,你可以试着使其中的部分内存失效(这通常可通过主板上的开关来做到),然后让 LIM EMS 4.0 扩充内存卡用扩充内存来填充失效的常规内存区。

这样,就创建了所谓的“可映射的常规内存”。这种内存就像是一个巨大的页面:EMM 设备驱动程序能够将所有回填内存换进或换出扩充内存。从理论上讲,LIM EMS 4.0 允许 EMM 设备驱动程序与扩充内存交换 8086 的整个 1 MB 的地址范围。

然而,实际上大多数 EMS 4.0 内存卡并不是都完全遵从 EMS 4.0 标准。比如,Intel 的 Above Board 只能回填 256 KB 到 640 KB 之间的常规内存。

注意 '386 微处理器具备强大的内存映象功能。8088/8086 和 80286 系统需要与 LIM EMS 兼容的硬件和软件,而 '386 则只需通过使用扩充内存仿真器,就能够在扩展内存中模拟扩充内存。有关此主题的内容将在第 5 章和第 9 章中讨论。

什么是 Shadow RAM?

Shadow RAM(影像 RAM)是指当计算机自举时,一种将视频和系统 BIOS 拷贝到重新分配的未用 RAM 区域,或“映射”到上位内存的技术。因为 RAM 比 ROM 更快,故此举能增强整个系统的性能。和高速缓冲内存(cache memory)一样,Shadow RAM 也不能被加进计算机中,必须购买一个已经装好 Shadow RAM 的系统。

如果想买一台带影像 RAM 的计算机,得保证影像 RAM 能被禁止。这种能被禁止的特色对于某些应用软件解决内存冲突来说是很重要的。

2.3 扩展内存

什么是扩展内存?扩展内存基本上就是指基于 80386 或 '386 微处理器的 PC 机中 1

MB 标志之上以及之后的内存。记住：之上和之后。所谓之上，是指 MS-DOS 所能管理的内存之上；所谓之后，是指大多数 MS-DOS 应用程序所能达到的内存之后。

80286 能访问至多 16 MB 的 RAM，而 '386 能访问多达 4096 MB 的内存。在所有这些系统中，MS-DOS 5.0 之前的 MS-DOS 版本都将 80286 及其后继的微处理器当作一个带 1 MB 内存的快速 8088 处理器。该 1 MB 标志之上的所有内存都被称为扩展内存。图 2.3 给出了典型的带扩展内存、基于 80286 或 '386 机器的内存映象。

扩展内存的最大问题就是它处在 8088 的地址空间之上，因此，MS-DOS 是不能直接使用扩展内存的。

为了使用扩展内存，程序必须将微处理器转换到保护模式，在退出前再将微处理器转换回实模式。（参见本章的下一部分：“实模式和保护模式”。）采用这一技术的第一个程序被用于 RAM 盘。

因为扩充内存的出现比扩展内存早，所以设计成使用扩充内存的程序比用扩展内存的要多。不过，这种局面正在改变。AutoCAD 和 Lotus 1-2-3 release 3. x 都能使用扩展内存，此外，Microsoft Windows 的增强模式也开启了数兆字节的扩展内存之门。

2.3.1 实模式和保护模式

要理解扩展内存，必须深入了解 80286 和 '386 处理器的性能。这些微处理器实际上都有两种操作模式，即实模式和保护模式。

- 在实模式下，80286 的操作方式和 8088 完全一样，它只能访问 1 MB RAM，扩展内存不能用来运行程序。

- 80286 在保护模式中得以解放。它能够存取多达 16 MB 的 RAM，并且可以在任何内存区中运行程序。扩展内存对于微处理器已经大大开放。

'386 微处理器和 80286 一样，也有实模式和保护模式。实模式中，'386 就像是一个超高速的 8088；保护模式下，'386 则能利用到多达 4096 MB 的 RAM 空间来运行程序、存储信息。此外，当运行支持保护模式的操作系统时，'386 处理器有一种被称为“虚拟 8086”，或“v-86”的模式。在该模式下，操作系统可以运行多个程序，其中的每一个程序都认为自己运行在一个独立的 8088 微处理器上，并且，只使用该处理器自身的 1 MB 地址空间。这就是 DESQview/386 使用 MS-DOS 应用程序多任务化的方式。

保护模式听起来就像是 RAM 饥渴者的美梦：有大量内存，并且微处理器能充分利用这些内存来发挥出威力。然而，这只是一个梦想。因为 MS-DOS 是与 8088 紧密相关的，它

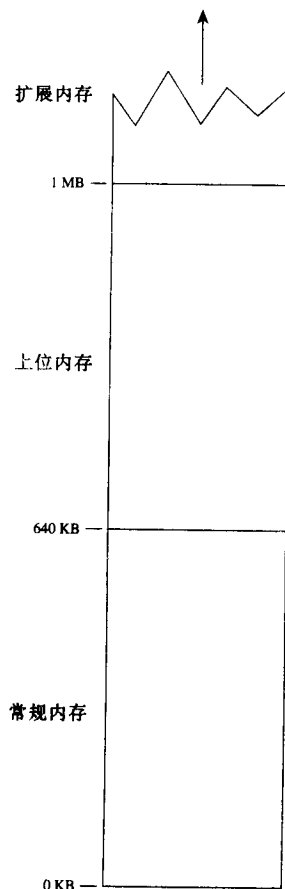


图 2.3 带有扩展内存的 80286 或 '386 计算机的内存映象图

是一个只支持实模式的操作系统。MS-DOS 不能运行于保护模式下,并且,它也不能在扩展内存中运行程序。

2.3.2 在 MS-DOS 下使用扩展内存

尽管迄今为止,扩展内存实际上仍是在 MS-DOS 的管辖之外,然而使用扩展内存的几种潜在优势还是存在的:

- 有些应用程序能够在扩展内存中运行,并且,它们将使用扩展内存。这些应用程序就是所谓的扩展的 DOS 应用程序。严格地说来,这些程序只是将微处理器从实模式转换到 MS-DOS 所不能及的保护模式而已。程序运行时,它们利用微处理器的强大功能,使用了机器中的所有扩展内存;程序退出,返回到 MS-DOS 时,它们则重新将微处理器置回实模式。这类程序中,最著名的有 Windows、Lotus 1-2-3 3.x 版、以及某些版本的 AutoCAD 软件。

- 从 MS-DOS 3.2 版开始,扩展内存就可开始用于内存存储。用户可以在扩展内存中设置 RAM 盘、磁盘高速缓存以及打印缓冲区。(详细内容将在第 7 章讲述。)只要程序不在扩展内存中运行,MS-DOS 就能将该内存作为存储区而访问它。

- 扩展内存可以用来模拟扩充内存——但是只能在基于 '386 的机器上使用,并且要有合适的软件(扩充内存仿真器)支持。

为了实现这些技术,Lotus、Intel、Ast Research 以及 Microsoft 公司建立了 MS-DOS 下扩展内存的使用标准,即扩展内存规范 XMS。

XMS 不允许程序在扩展内存中运行,用户仍然被限制于上面所描述的三种功能。不过,XMS 确实建立起了能在 MS-DOS 环境下提供扩展内存精巧、协作使用方法的标准。

2.4 MS-DOS 的综合解决方案

MS-DOS 5 版敞开了 PC 机使用内存的新大门,特别是对于基于 80286 和 '386 的系统来说更是如此。MS-DOS 6 继承了这一传统。随着新方法的出现,同时也出现了用于描述内存的新术语。

2.4.1 高端内存区(High Memory Area)

8088/8086 微处理器只能寻址 1 MB 内存空间,任何寻址 1 MB 以上内存的企图都将导致处理器回绕(wrap around),即退回到内存 0 位置。(就象很早以前的小行星视频游戏,当船驶出屏幕的顶端时,小船将在屏幕的底部重新出现。)

80286 和 '386 微处理器也要回绕到内存 0 位置处,不过,它们还能将超出 1 MB 的字节映象到扩展内存的第一个 64 KB 段。这些额外的 65520 字节即所谓的高端内存区,或 HMA。参见图 2.4。尽管 HMA 实际上比 64 KB 少 16 字节,我们还是假定 HMA 向 MS-DOS 提供了额外的 64 KB 内存。

在基于 80286 或 '386 的机器上,MS-DOS 可以使用 HMA 这块额外的内存区。MS-DOS 能确切地“看见”HMA 所处的位置,并且能直接访问它,而不用将微处理器转换成保